

```

1  TTCGCCTGCG  GGCCGGCACT  GCTCACCTCT  CGTCCAGGGA  CATGACGGGC
51  ACGCCAGGCG  CCGTTGCCAC  CCGGGATGGC  GAGGCCCCCG  AGCGCTCCCC
101  GCCCTGCAGT  CCGAGCTACG  ACCTCACGGG  CAAGGTGATG  CTTCTGGGAG
151  ACACAGGCGT  CGGCAAAACA  TGTTCCTGA  TCCAATTCAA  AGACGGGGCC
201  TTCCTGTCCG  GAACCTTCAT  AGCCACCGTC  GGCATAGACT  TCAGGAACAA
251  GGTGGTGAAT  GTGGATGGCG  TGAGAGTGAA  GCTGCAGATC  TGGGACACCG
301  CTGGGCAGGA  ACGGTTCGGA  AGCGTCACCC  ATGCTTATTA  CAGAGATGCT
351  CAGGCCTTGC  TTCTGCTGTA  TGACATCACC  AACAAATCTT  CTTTCGACAA
401  CATCAGGGCC  TGGCTCACTG  AGATTCATGA  GTATGCCAG  AGGGACGTGG
451  TGATCATGCT  GCTAGGCAAC  AAGGCGGATA  TGAGCAGCGA  AAGAGTGATC
501  CGTTCCGAAG  ACGGAGAGAC  CTTGGCCAGG  GAGTACGGTG  TTCCCTTCCT
551  GGAGACCAGC  GCCAAGACTG  GCATGAATGT  GGAGTAGCC  TTTCTGGCCA
601  TCGCCAAGGA  ACTGAAATAC  CGGGCCGGGC  ATCAGGCGGA  TGAGCCAGC
651  TTCCAGATCC  GAGACTATGT  AGAGTCCCAG  AAGAAGCGCT  CCAGCTGCTG
701  CTCCTTCATG  TGAATCCCAG  GGGGCAGAGA  GGAGGCTCTG  GAGGCACACA
751  GGATGCAGCC  TTCCCCTCC  CAGGCCTGGC  TTATTCCAAG  AGGCTGAGCC
801  AATGGGGAGA  AAGATGGAGG  ACTCACTGCA  CAGCCGCTTC  CTAGCAGGGA
851  GCTATACTCC  AACTCCTACT  TGAGTTCCTG  CGGTCTCCCC  GCATCCACAG
901  GGAGGGTAAA  ACACCTAGCT  TTTATTTTAA  TAGTACATAA  TTTAATACCA
951  AAAAAGGCGC  CTGGATCCCC  AAAAAACCGA  GGCTGGGAGC  TAGTGGCCCT
1001  TTTGCTTTCT  AGGACTTGGG  GGGCCGGCCC  TCCCTCCTAA  GCATAACAAA
1051  GGTGGTGTG  CTCAGCTCA  GCCCCAGGGG  ACACAGATGC  ACTTGGGGG
1101  TGAGGGCAGG  TAATGACTCC  ATCGCACCC  CAGTTCAGCT  GGACAGAGGC
1151  TCAGGTGACC  CCAGCCTTCA  CTGTCTCCG  CTCTCCAGGA  GCTTATCTTC
1201  GCCCCATCTC  CCAAATAAGT  GGGCCCTTGT  GCTGTGAGGA  AGACCAAAGC
1251  CTCAGGGAAG  ATAAGAGATA  TGGAGATGGG  AGGGGGAGGA  CAAGGGGCAG
1301  AGAGTAGGGT  CTAGCTGGCT  ATCTCTGGCC  TTACTAACAC  CCCCTGGAG
1351  GCATGCCCT  TTTCTCCAGC  ACACAAGCAC  ATTGGGGCAC  CTGGAAATAT
1401  TGGTTCAGG  CTCCTGTTCT  CTGGACTTCA  GATCCTGGG  GAGCCCCTCC
1451  CCCCCCTGAA  TCCCTGGCTT  AGCTACCTTC  CTGCCTGTGC  ACCTAAAAAC
1501  CTCAGGTCAG  AACTAGGAAA  AGAGTTTGT  TTTTATTTT  TTGAAATGGA
1551  GTCTCGTTCT  GTCGCCAGG  CTGAGGTGCA  GTAGTGCAAT  CTCCGCTCAC
1601  TACAACCTCC  ACTCCCTGGG  GCTCAAGCGA  TCCTCCCACC  TCAGCCGCCG
1651  AAGTAGCTGG  GACTATAGGT  GTGTACCATC  ACACCTGGCT  AATTTTGTGTA
1701  TTTTTGTAG  ACACAGGGT  TCGCCATGTT  GCCCAGGCTG  GTCTTGAATT
1751  CCTGAGCTCA  AGCAACCTGC  CGGCCTCGGC  CTCCCAAAGT  ACTGGGATTA
1801  CACGCAGAAG  GCACCATGCC  CAGGCTAGAT  GTGTCTTATC  CCAATCCTTT
1851  GGCAGGCATG  CAGCTCCACA  GCGGATTTCT  TCAAGCAGCT  GAAGTGTTTA
1901  GCCCTCCTGG  GTTAAGAGCC  AGATAAGGAG  AAATCCCTTT  CCTAGGTTTG
1951  GAATGTGTTG  TGAAAAAAA  GAGAAATCCC  TGGCTCCTGG  AGCTGGTGGG
2001  AGACAAGATT  AAGCAAACCT  CCCCTGACAT  GTATCCCTTT  GACCCCAAGC
2051  TCTGCCCTCT  CCCTGACCAC  CCATGCCCTT  TCCTTTAAGT  TCTCAAACAG
2101  ATACCAGGGC  CTAACTGCT  TTACCTCCC  TCCTACTGAG  TCAGGTTAGG
2151  TGGTGGGAGG  TCACCCATTT  CCGAGTTAAA  CCAATGCAAT  ATGAGTAAAA
2201  CAAAGTCATG  TGGGTATGTC  TGGGGTAGAG  AGAGGGGTAG  CAAGTTCATG
2251  TGTCTCCTT  GGTCACATAT  CTCCCAAAGC  TCTGATCCCT  GCCATGGGAA
2301  GTGGACAGGA  AACATGAGGT  CATGACCTGC  AGGCATCTTT  ACTGCAGCTC
2351  TGCCGGCCTG  GAGGGGGAGA  GGGGGAGGAA  GAAGTATGCG  CTGCACATTT
2401  CTGAGGCTAC  TGCAATTGCT  TTCAAGGCAG  AAATCTTGCT  CTGAGCAGTC
2451  AGCGGCTCCA  GTTTGGGCC  GATAAGGAAG  TTCTCCGTGG  CCTCCCTCAG
2501  GCAGAGCAGG  GAGGAGGCTG  ACATTGCCAG  TCTCTTCTGG  GGCCCAAGGC
2551  AGGTTGCAGG  AGATCCAATC  CCATAGACAG  CTCTGGGCCT  CTTGCATTTG
2601  AGTTTTTCAG  AATTAAACTG  CAGTATTTT  GAAAGCAAAA  AAAAAAAAAA
2651  AAAAAAAAAA  AAAAAAAAAA  AAAA  (SEQ ID NO:1)

```

FEATURES:

5'UTR: 1-41
Start Codon: 42
Stop Codon: 711
3'UTR: 714

Homologous proteins:

Top 10 BLAST Hits

	Score	E
CRA 103000001517087 /altid=gi 10946770 /def=ref NP_067386.1 RA...	425	e-117
CRA 1000682330460 /altid=gi 7657492 /def=ref NP_055168.1 RAB26...	297	4e-79
CRA 18000004977238 /altid=gi 1710022 /def=sp P51156 RB26_RAT RA...	294	3e-78
CRA 18000005013109 /altid=gi 1083775 /def=pir JC2528 GTP-bindi...	293	7e-78
CRA 89000000198627 /altid=gi 7296421 /def=gb AAF51708.1 (AE003...	273	9e-72
CRA 18000005076419 /altid=gi 7438397 /def=pir T15123 hypotheti...	207	4e-52
CRA 18000004912300 /altid=gi 134236 /def=sp P20791 SAS2_DICDI G...	203	7e-51
CRA 98000043536338 /altid=gi 12963499 /def=ref NP_075615.1 cel...	203	9e-51
CRA 18000004929618 /altid=gi 131798 /def=sp P24407 RAB8_HUMAN R...	202	1e-50
CRA 18000004952869 /altid=gi 131848 /def=sp P22128 RAB8_DISOM R...	202	2e-50
CRA 18000005221564 /altid=gi 4586580 /def=dbj BAA76422.1 (AB02...	202	2e-50

BLAST dbEST hits:

	Score	E
gi 13033710 /dataset=dbest /taxon=960...	1318	0.0
gi 12785775 /dataset=dbest /taxon=960...	1316	0.0
gi 12904236 /dataset=dbest /taxon=960...	1035	0.0
gi 9093496 /dataset=dbest /taxon=9606...	694	0.0

EXPRESSION INFORMATION FOR MODULATORY USE:

library source:

From BLAST dbEST hits:

gi|13033710 prostate
gi|12785775 brain
gi|12904236 T cells from T cell leukemia
gi|9093496 leukopheresis

From tissue screening panels:

leukocyte

FIGURE 1, page 2 of 2

1 MTGTPGAVAT RDGEAPERSP PCSPSYDLTG KVMLLGDTGV GKTCFLIQFK
 51 DGAFLSGTFI ATVGIDFRNK VVTVDGVRVK LQIWDTAGQE RFRSVTHAYY
 101 RDAQALLLLY DITNKSSFDN IRAWLTEIHE YAQRDVVIML LGNKADMSSE
 151 RVIRSEDGET LAREYGVVFL ETSAKTGMNV ELAFLAIAKE LKYRAGHQAD
 201 EPSFQIRDYV ESQKKRSSCC SFM (SEQ ID NO:2)

FEATURES:

Functional domains and key regions:

[1] PDOC00001 PS00001 ASN_GLYCOSYLATION
 N-glycosylation site

114-117 NKSS

[2] PDOC00004 PS00004 CAMP_PHOSPHO_SITE
 CAMP- and cGMP-dependent protein kinase phosphorylation site

Number of matches: 2

1 214-217 KKRS
 2 215-218 KRSS

[3] PDOC00005 PS00005 PKC_PHOSPHO_SITE
 Protein kinase C phosphorylation site

Number of matches: 5

1 29-31 TGK
 2 113-115 TNK
 3 149-151 SER
 4 173-175 SAK
 5 212-214 SQK

[4] PDOC00006 PS00006 CK2_PHOSPHO_SITE
 Casein kinase II phosphorylation site

116-119 SSFD

[5] PDOC00008 PS00008 MYRISTYL
 N-myristoylation site

Number of matches: 5

1 3-8 GTPGAV
 2 6-11 GAVATR
 3 39-44 GVGKTC
 4 52-57 GAFLSG
 5 57-62 GTFIAT

[6] PDOC00017 PS00017 ATP_GTP_A
 ATP/GTP-binding site motif A (P-loop)

36-43 GDTGVGKT

[7] PDOC00579 PS00675 SIGMA54_INTERACT_1
 Sigma-54 interaction domain ATP-binding region A signature

32-45 VMLLGDTGVGKTCTF

Membrane spanning structure and domains:

Helix	Begin	End	Score	Certainty
1	48	68	0.715	Putative

BLAST Alignment to Top Hit:

```
>CRA|103000001517087 /altid=gi|10946770 /def=ref|NP_067386.1| RAB37,
    member of RAS oncogene family; GTPase Rab37 [Mus
    musculus] /org=Mus musculus /taxon=10090 /dataset=nraa
    /length=223
Length = 223
```

```
Score = 425 bits (1081), Expect = e-117
Identities = 209/223 (93%), Positives = 215/223 (95%)
Frame = +3
```

```
Query: 42  MTGTPGAVATRDGEAPERSPPCSPSYDLTGKVMLLGDTGVGKTCFLIQFKDGAFLSGTFI 221
          MTGTPGA      DGEAPERSPP SP+YDLTGKVMLLGD+GVGKTCFLIQFKDGAFLSGTFI
Sbjct: 1  MTGTPGAATAGDGEAPERSPPFSPNYDLTGKVMLLGDSGVGKTCFLIQFKDGAFLSGTFI 60
```

```
Query: 222 ATVGIDFRNKVVTVDGVVRVQLQIWDTAGQERFRSVTHAYYRDAQALLLLYDITNKSSFDN 401
          ATVGIDFRNKVVTVDG RVKLQIWDTAGQERFRSVTHAYYRDAQALLLLYDITN+SSFDN
Sbjct: 61  ATVGIDFRNKVVTVDGARVQLQIWDTAGQERFRSVTHAYYRDAQALLLLYDITNQSSFDN 120
```

```
Query: 402 IRAWLTEIHEYAQRDQVIMLLGNKADMSSERVIRSEDGETLAREYGVPFLETSAKTGMNV 581
          IRAWLTEIHEYAQRDQVIMLLGNKAD+SSERVIRSEDGETLAREYGVPF+ETSAKTGMNV
Sbjct: 121 IRAWLTEIHEYAQRDQVIMLLGNKADVSSERVIRSEDGETLAREYGVPFMETSAKTGMNV 180
```

```
Query: 582 ELAFLAIAKELKYRAGHQADEPSFQIRDYVESQKKRSSCCSF 710
          ELAFLAIAKELKYRAG Q DEPSFQIRDYVESQKKRSSCCSF+
Sbjct: 181 ELAFLAIAKELKYRAGRQPDEPSFQIRDYVESQKKRSSCCSFV 223 (SEQ ID NO:4)
```

Hmmer search results (Pfam):

Model	Description	Score	E-value	N
PF00071	Ras family	306.9	8.4e-90	1
CE00060	CE00060 rab_ras_like	213.3	3.7e-60	1
PF01142	Uncharacterized protein family UPF0024	2.6	3.4	1

Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
CE00060	1/1	31	191 ..	25	193 ..	213.3	3.7e-60
PF01142	1/1	185	201 ..	444	462 .]	2.6	3.4
PF00071	1/1	31	223 .]	1	198 []	306.9	8.4e-90

1 AGGGGAGAGA AAAGACCGCA TACCAGGCCA GGTGCGGTGG CTCACGCTTG
 51 TAATCCCAGC AATTTGGAAG GCCAAGGCAG GCGTATCGCC TGAGGTCAGC
 101 AGTTCCAAAC CAGCCTGTCC AACATGGTGA AGTTCTCTAC TAAGAATACA
 151 AAAATTACCC AGGCGTGGTG GCGTGACCT GTAGTCCCAG CTGCTCCAGA
 201 GGCTGAGGCA GGAGAATTGC TTGAACCTGG GAGGCAGAGG CTGCAATGCG
 251 CCAAGATCCC GCCACTGCAC TCCAGCCTGG GCGACAGAGT GAGACTCCGT
 301 CTCCGGGAGC CCACGGCATT GAGCAAACCT CGGCATTATT TGCAGCAAGA
 351 GCCTCTGGCA TCCAAATAGC AACCAACACC ACGCCTCTGT AGTGTGCTGC
 401 GCAGCCTCCA CACTCCAGTC TGAGGCTCCC TGTTTGAGTC CCGCCCTATG
 451 CCCAGCTGAG GTTATAGCAC GCTCACCTCC AGAAGAGGTA ACCCAAGCTC
 501 TTTACTCTAC TGGAGATCAC CTCTGTCCCC ACTCTGGGCG CTTCTCCCAG
 551 CTGACAGAAA ATACCTCCAG CTGATGTCAG AAAATACAGG GCTGGAGGCT
 601 GCGGTACAAA GTCAGTCCCC ACAGGCCTAT GGTGGCCCAT AAGCCACGTC
 651 TACCCCTGCT CCTCACCTCC ACACCTAAGT TAAGAATTGC AGGCCGGGCG
 701 CAGTGGCTCA CGCCTGTAAT CCCAGCACTT TGGGAGGCTG AGGTGGGCGG
 751 ACCGCTGAG GTCAGGAATT TGAGACCAGC TTGGCCAACA TGGCAAAACC
 801 CCGTCTCTAC TAAAAATACA AAAAGAAAAA ATAGCCGGGC CTGATGTGCG
 851 GCACCTGTAA TCCCAGCTAC TCCGGGAGAC TGAGGCGGGA GTATAGCTTG
 901 AACCCTGGAA GCAAAGGTTG CAGTGAGGCG AGATCGCACC ACTGCACTCC
 951 AGGCTGGGCG ACAGAGTGAG ACTCTGTCTG AAAAAAAAAA AAAGTGCAGG
 1001 TACCCCTCTC CAGCTCTCCC CTCCCTACAC ATCCCTCAAA CCGTCCCGCT
 1051 GTAATGCACC CGCCCTGTTC CTGGGTAAC TGAAGCTGCT TATAGAATGT
 1101 GGAGATGGGG GTAATTGAAA GGTGCGCCCA GGCCACAGAG CCCCTGAGCT
 1151 CTGCTACCGG CAACCCAGC TGCCTCCCC ACTCTCTGTC ACCAGGAGCT
 1201 GCGGGGTGCC TGGGATATCC TGGCAGCTCT GCTCAAAATG ATCTACGACT
 1251 TCATGAATTT ATTTGGCTCC TCCTCGGGGC CAGGGTGAGT GTCATGGGTT
 1301 AATAAGGCCG GCCCCGCTT CAGGAGCGGT CCACTGGGAG ATGTGTGCTG
 1351 CGCAGCCCTC TTGCGAAAGC TCTCCCTGG TGGGACATTC TGGGCACAAC
 1401 CAACAGGCCG GGGGAAATGA GAGGTGATCC ATACTAAAGG GTCAAAGTCC
 1451 CCGACCCGAG CAGAGGCCCC AAAACACCGC AGCGTACATG TGCTGCAAGG
 1501 CGAGTACGGG TTGGTAAACA AAACATATAT CAGATGAGCT CGGGCCGGGT
 1551 GACTTAACAG ATGAGGAAGT GTCTCGGGGC CATCGGCGGA GGCGCAGCCC
 1601 AGGGGTCCCC AGCTCCCCGC CTCGCCACCT GGGGACAGCC CACGGCCCCG
 1651 GGCTCGGGCG CCGCTGCTG TCGCCGTGCG CAGCGACTAC GGGAACTCTT
 1701 CCGCAGCAGA CGGGGTCCCC GCGGCCCGCT CCCCAGGGG CAAGCAAGCG
 1751 ACCACAGGGG ACCGGTCCCC GGGCTGGATG TGCTCATGT CCGAAGCGCA
 1801 CGGAGCCGAG CCGGTGTTGC TCAGGGAGGC TGCCCGCCCC TTCACGCAGA
 1851 CCCTGCGGCT CTGCGTGCCC TCAGGGAACA GCAAGGTCCG AGCCGGTGTC
 1901 GTCGAGGGGG CGACGGGACG GAGGGAGGAG CCTGAGGGGT CCCGGTCGAG
 1951 GGAGGGGAGG AGTGGGCGGG GCGGGGTGG GGGCCGTTCC CGCGCTCTCC
 2001 TTCGCCTGCG GGCCGGCACT GCTCACCTCT CGTCCAGGGA CATGACGGGC
 2051 ACGCCAGGCG CCGTTGCCAC CCGGATGGC GAGGCCCCCG AGCGCTCCCC
 2101 GCCCTGCACT CCGAGCTACG ACCTCACGGG CAAGGTGGGT GGGCCTCTTC
 2151 CGTGAGACCC CCGCCCTCCT CGCGCCTAGC CCCTTCCTGG CTGCGTCTGG
 2201 GTTGGACTCA GCCCTTCCCC CAGGCAGCTG CGTCTCCCAG AGGAGGGAGG
 2251 GAGAGAGGGT CAGGACACAG CCTCTGGGGC CGTCCCAAGC TCTAGGTGTC
 2301 TCTGCTGGCT TGGTGGGGGC GGGTCGCGGA AGATCGCAAA AACTGAGTGA
 2351 TCCCCCGGCC GGCCCCAACT CAGTTCTCTT CTGCCACACT CTGGCAAATA
 2401 TGAGCCCCCG GGAGCCCATG CTCTTGGTG AGGGTTAAGC GCGCAACTCT
 2451 CGGGGCTCAG GCTGGGAAGG GCTGGGAGAT GGGGACCGAA CGGAGACTCG
 2501 GAGAGGACGT CCCCTGCTGG CAGAGGAACT GGCGTTAATG CCATTTTCCG
 2551 AGCTAAGCTC TTAGTTGAGA TCTGACATCC AGGTTTAAGG CCTGATGTCC
 2601 CCCAGCTGCT CCCCTCCCAT TCCACCCGCT GGAGGCACTG CCTCCCACCT
 2651 TCCTCCCTGC AGTCGGAAGC CGCTCCTCCC AGAAGGATGT TGCCAGCCGG
 2701 CCTGCAGGTC ACTTGGGAAT TTTTCGAACC TGAGAAAGAT TTCAGTGGTT
 2751 GGTCTTTTCG ATCCCGCACT TGAGAGAGCT CCAGGGCTGC TCTCTGGGGC
 2801 TTGCTCCCTC TACAGGGGTG TCCTGTATGG AAACAGGTAG GGACAGCAGT
 2851 GGACTGGTCT GTCGCCTTCC ATCTGTGTCC TTGGAGTGAG CGGGTACCAG
 2901 AAACGAAAG AACTGCTGAG GGAGCCTAGA GCTTCCACTC TTCCTCTGCA
 2951 GGGTTGGGGA TGGAGTGAGG GCTGTCCTGG ATTCCGCTGC ATGGCCTTGA
 3001 AGGAGACCTG CCTCTCTCTG GGCTCGGTT TCCTCCCCGA CACCAGGGCT
 3051 CACCCTTGCT GGGAGCCTCA GCCTCCACCC CAGTGTTTCG GGGGAAGCCA
 3101 CCCTGCAAGT CATCCGCCCA GAGCCGTTGA GATAGGCGTC CTGTGTGGGC

FIGURE 3, page 1 of 7

3151 TTGTGGCAGG AAATGGGCCC CTGCACCCTC GGAGAGGAGG AGCTGCTGTT
3201 GGCCAGGCCC CAGGCTGAGG GGGACTGCCT GACCTTGTTG CCCTGCAAAC
3251 CAGCTGGGTT GTTTGCCTAG GAGGTGGCCA GGCTAGGCAG CTGTTTGTGT
3301 TTGGTGAAT CACCGAGCTG GGTGGGTAGC TGGCATCGTT TGCTCAAGGC
3351 AGCTGTGATC TGTAAGTAC ACAAAGACTG GCCCTCCCTC CCTCCTTCCT
3401 GCTCCAGGGC TGGGACCCAG GAGCCAGGGA GGAGTGCAGG CTCCAGAAAG
3451 CTCCTATCCC CCACCCCTTC ATCTGTTCCT TGGCCAAGCG GCATTGGCCG
3501 GAGAGTTGGT CCCCAGCCTC CCCGGGCTG CCCCAGGGGA GTGAGTCCAG
3551 GACCTCTGA GAAAGCCTGG CAGGAGCTCC TTGGACCAGA CTAGGGGTGA
3601 TGTGGCCAC AGGCAGACAG TTCCACCTT GGGCCACTCT TCCCTGGGTC
3651 TTAGGTGATT CACCACGATG ATGGGCCCTA GCCATTAACA GACTCTAGAA
3701 ATACCTCAA GACATTATCC CTCCTCCTTC TACCCACTAT GGAAACCATG
3751 CCACAGAAAG GTTAAGGAAT CTTCTAAAG TCACACAGTA GGCCATTTAC
3801 AAATCAAGAC CCATCCTTCA TACCCCTTCT GCTCAGCCAC CCCTGCCTCT
3851 CCACCAGAGT TAACTAATGC CAGTACCCCA TGCCCAACA AGGAATGCCT
3901 TTGGGTCCA CTGTCAATTT CAGAGCCTCA AAAATAATT AAACCTAGTC
3951 CCTGCTTAAC ACATTAAGCC ACCTAACCAG CAGCTGGGAA ATCCAGCAT
4001 TGGATCTAGA CCCCTGTTAT CCAAGATTGG AGAACAGTGG GACAAAGTGC
4051 TCCTCTCCAC CATTCCTGCG TGTCCCTGGG GAAGATGAGC AGAGCAGAGC
4101 CAGACAGTAA AGGAGAGGGC CACGCCCTC CCACAGGTTA CCTCCTTGGT
4151 ACTCCTGCCC GCACTACCCA CAGCAACCCC GGGATGCCGA TCTGCAGCCA
4201 CATGTCCCCT GTGGGAGGTT TCTGCTGAAA GAACTTCCAA CTACACATCT
4251 CCCCCTTCA GTATAAATTT CAACCTTCCC TAATTCATGC AACCTTTTTT
4301 TTTTTTTTTT TTTTTTGAGA CAGAGTGTG CTCTGTCAAC GAGGCTGGAG
4351 TTCAGTGATG CAATCTCGGC TCACTGCAAC CTCTACCTCC TGGGTTCAAG
4401 CTATTCTCCT GTCTCCGCT CCCAAGTAAC TGGGACTACA GGCGTGTGCC
4451 ACCACTCCTG GCTAGTTTTT TGTATTTTTA GTAGAGATGG GGTTCACCT
4501 TGTGGTCTAG GCTGGTCTCA AACTCCCAAC TCAGGTGATC CGTCCACTTG
4551 GGCACCAAAA ATGNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN
4601 NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN
4651 NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN
4701 NNNNNNNNN NNNNNNNNN TTCAAGTACC AGCCTGGCCA ACATGGTAGA
4751 AACCCGCTCT CTAATAAAAA TAAAAATTA GCCAGGCGAG GTGGTGCATG
4801 CCTATAATCC CAGCTACTCA GGTAGGCTGA GGCAGGAGAA TCATTTAAAC
4851 CTGGGAGGTG GAGGTTGTGG TGAGCCAAGA TCTGCCATT GCACTCCAGC
4901 CTGGGCAACA AGAGCAAAAC TCCGTCTCAA AAAAAAAG AAAGAAAGAA
4951 AGAAAGAAAC TTCCAAATAA ATGTTGTGAC ACAAAAAA AAACCAAAAC
5001 AATATTCATT ATAGAGTATG CAAATGACCA TGCCCCACCC CCAGCAGATT
5051 CTGATAGACT CCCTTGGGTG GGAATCCTTG TCCAATATAT TGACACTTCC
5101 CTTTCTGTG AGTATAGCCC AGCCATGCG TGTACTCACG AGCGGACGAT
5151 GGATGACACA AGTACACAGA GGGACGGAAT CCCTGCATGG TGTGGCTATG
5201 GGCAAAATGT GCCACTGTCT AGATTGTGCA AATGTGGTGG TTCTCTGGG
5251 CCACAGAGCA CACTTGGGGA CCTGTTCATG GTGAGGTCTC AACTCCGGCC
5301 TCTAGGAAC TGAATGAGGA CAGGAGGGTC AGAGGGAGAG CCTAGGAGGC
5351 TGAGCCAAGG AGCGTGGAGA GGAGAGACAG GGTGAAGGTG GCGGCTGGCT
5401 TTCTGGAAGC AGGTGGCCTT TGGTGGGTC AGCATTCGTG CCAGCCCCCT
5451 CTTCTCTGAT CCTCTCCATG TGTCTCTCTC CTGGAATCCC AGAAGCTGCC
5501 CTTGACTCCC CATTAAGTGC CTCTGCCCCC ACCCCCTAGG TGATGCTTCT
5551 GGGAGACACA GCGTGGGCA AAACATGTTT CCTGATCCAA TTCAAAGACG
5601 GGGCTTCCT GTCCGGAACC TTCATAGCCA CCGTCGGCAT AGACTTCAGG
5651 GTGAGGTGGC TGCAGGCACT TGCTTCCAGC AGAGAGCCAG GGCTGTGGCT
5701 CAGGCATGGG GGGGTGCCC CCACCTTGCT CACCCTGGCT CCCAGGGACT
5751 CCCGAGGCTC ATGCCTGGAG GGCACACAAC CCGTCCCCC AAGACCACAG
5801 AGGTGGCCGG GTCAAAGGAG ACTGGGCAAG GTTGGCTCCT TGCCCAACTA
5851 TAGGATGCAA AAAAATGAGA CTGAGTCTTC GATTCCAGCT CCATTCTTGG
5901 GGGACTTCTC CCAAGCAGAG CAGCCGAGG CACGGCATAA GCTGAATATC
5951 TTGGCCCA GAGCCCTGC TCATTGCTCT CCTACCTGGG CCCCTTTGGA
6001 AAGGCCTCAA AGGTCAATCA GTCTTTCTGG AGTTCCAGA AAGCACAGCC
6051 CTGCACTGGG TTTAAGAGCT GGGCTTGGG CAGGCATGGT GGCTCTTGCC
6101 TGTATTCCCA GCACTTTGGG AGGCCGAAGC GGTGAGATCA CAAGGTGAGG
6151 AGTTTGAGAC CAGCCTGGCC AACATGGTGA AACCCCGTCT CTAATAAAAA
6201 TACAAAAATT AGCCAGGTGT AGTGGCACGC TCCTGCAGTC CCAGCTACTC
6251 GGGAGGCTGA GGCAGGAGAA TCGCTCAAAT CCGGTGGTG GAGGTTGCAG

FIGURE 3, page 2 of 7

6301 TGAGCTGAGA TCGCGCCACT GCACTCCAGC CTGGGCAACA AAGTGAGACT
6351 GCGTCTCAGA AAAAAAAAAA AAAAAAGAGC TGGGCTGGCC ATGTTGGGAG
6401 ACAGCAGCTC ACCAGGGACC CTCCCTCTCA CCTTGACGAC TCCATCTTAC
6451 AAATCTGCAT CAGGGATGCT AGACGCTGCA CACCTGAAGT GTTCAATAGA
6501 GAAAAGGTCT CACCCTGGCA GGTGGGGCTC TACAGCTTCA AGCAGGCAGA
6551 AAGCGAACAC TTCCTTCACT AGAGAATTAG TGGGCAGCTA AAGAAAAGGT
6601 GCTGCTGCAG ATGTAGCCTC AGGTCCCCAG GATGCAGGCA AACACCCCAT
6651 CTCCAGGGGC TCGGTACAG TCCCAAGGCT AGGCTCCAGG AGAGGGAGAC
6701 CGAAGTGGGG AAAGGGCAGG GCCTCCAGCA GCAACCAGCC CTCCAGCCCT
6751 GGGCTGCTG ATCCCTGGAG AGAGCCAGGA TGTTTCTCAG GCTCCTCTTG
6801 CCCTGCTGTT GTGAGAAAGC AGTTACAGTC CTCAGAAGGG ACGACTCCAC
6851 AGTGGAGGTG TCTGGGTATG GGGTTCCTGC TGCCCTGATG GTATGATCTG
6901 GCTGGAGACG GTTCTGGGGC TCACTGCACC CACTCTAGGC CTGGAGAGGG
6951 AACAAGAGAG GACGTCTGCA GAGCTGAGGA GCCACATGAC TCCTGCCCTC
7001 CCATCCTCTG CCTTTTCTC TTTCAGAAAC AGGTGGTGAC TGTGGATGGC
7051 GTGAGAGTGA AGCTGCAGGT GAGACCAGAG GCTGGAGTTG GGGAGGGAGG
7101 ATGGAGGACC TGCCCTTCCT TCTCACCCTG AACCACAGGA GGCTGCGAGC
7151 CCTGCCCTCC GCCTGGGGCA ATTTCTCTGT GGGCCACCGG GAGGAAATGG
7201 CTTTTGTTTA TTTGACATCT GCAGAAAAAG CAGTCCCAG GCACCCTCTC
7251 ATCTATGAAC AGCAGCTCCA AATGCCTTCA GACAAGCTTA GCCTCCATCC
7301 ATCTCCTCCC CAGTTGCCAG GGCTTTATCT GCTCTTAGGA GATTGGACAT
7351 CCCCACCCCC TGAGCTAGGG GAGAGGAGAA GATTCTTTT TTTCTTTTC
7401 TTTCTTTTT TTTTGTGAGA TGGAGTCTCG CTCTGTCGCC CAGGCTGGAG
7451 TGCAGTGGCA CAATCTCGGC TCACTGCAAC CTCTGCCTCC CAGGTTAAG
7501 AGATTCTCCT GCCTCAGCCT CCTGAGTAGC TGAGACTACA GGTGCATGCC
7551 ACCACACCTG GCTAATTTTT TGTATTTTGA GTAGAGACGG GGTTCCTG
7601 TGTTAGCCAG GATGGTCTGG ATCTCCTGAC CTCGTGATCC GCCTGCCTCG
7651 GCCTCCCAAA GTGCTGGGAT TACAGGTGTA AGCCACCGCG CTCGGCTGAG
7701 GAGATGATTT TGAACGAGCT TGAGAAATCA GTAAGTCTA CTGTCCAGGT
7751 CATTGGATGC TCAGGGGCTC ATGAGAACCT AAAGAAGAAA ACAGCCCCAC
7801 CTTCCACAG ATATCTCATA CAACAAAGCA GGCTGCTCC ACCCAGCACA
7851 TTCCTTGAC CTGCCTCCTT CTGACCATT CTCCATCCCA TCCCTTCCCA
7901 GATCTGGGAC ACCGCTGGGC AGGAACGGTT CCGAAGCGTC ACCCATGCTT
7951 ATTACAGAGA TGCTCAGGGT GAGTCCCTCG CACCTCCAA CCCCTACCCC
8001 AGCCCCTTGG TAGCATCCGT GCTGCTGCCT AAGTCCCCTC TGTGATCCTC
8051 TCCCCTCCAG CCTTGCTTCT GCTGTATGAC ATCACCACAA AATCTTCTTT
8101 CGACAACATC AGGTGAGTCT CTCCCTTCCC CTGACTCCCA CCCATAAGCA
8151 GCCAAGGCAA GGTCTATGCA GGCTGGGGTT GCTTCTTGCC CTGTGGAAG
8201 CGGGTGGAGC GTGGAGTCTT CTTGCCCTTCT GAAAAACACC TACTTGTGAC
8251 TCAGAAAGTCA TATCTGCTGC TTTGTATTTG GTGGCCATGT GGGCATGAAG
8301 GCCAAGCAGG CTGTTGTGAC CCTGTGCCAC CTGCATAGCC CTCACTGTGA
8351 TTCACGAGTG TGTTTCGTGA CAAAGTGTTT AGAACAGCCC CCACTCCACC
8401 CTGGATAATT ATCCACAGAG ACCAAGGGAA AAACACAACC AGAAAAGTCC
8451 ACACATACAT CCAGGGCAAG TTGCAAGAAA GTGACTCAGT CAGACAGAGT
8501 GAGTGGTTGT ATCTCACAA CCAAATATT ATAGAGACAA AAATTTGATA
8551 AATTCAAGCA CCAATTTTGT TCACGACATT GTATAGGTTT CATGAATCCC
8601 CTGACCTCAA GGACAGTTTG CTGATAAGCA AACTAGGAGA ATAAAACGTT
8651 TATATAGAAA GAGGAAAATC CATGGCACTC ATACTCCTAC CTCCAACCCC
8701 ATGCTCATGG CAGACATCAC TAATCAATCA CAGTACTTTT GATCACTGAA
8751 ACCCTTATGT GGTCTTAGAA TCTTTAACAG GACACTCCAA GAAATCACTG
8801 CTGACAGCCA ACTGATTTGT GAGATAAGGT CTCCATGCAT CTGGATCTTC
8851 CATAGAACTG ATAGTTGCAC AGCATAAAAT GGTGAGGGTG GGGCCATTGT
8901 GGGTTGAGCC ACCAAGGAAG GCCATCCAGG CCTGGATGGG CCAGAACAAA
8951 GGTACAGATG AGAGAACGCA CAGGGTATCG TGTTCAAGGT AGTGAGTAAC
9001 TGAGGATAGT CAAACGGAGC AGAAGAAGAA AGGGGCAGCA GGAGGAAGAG
9051 AATGCCAGTC TCGCAGGCC TCTCCACAG GCCTGGCTCA CTGAGATTCA
9101 TGAGTATGCC CAGAGGACG TGGTGATCAT GCTGCTAGGC AACAAAGTGA
9151 GTGGCTCCGG GGCAGGGTCA GCCCAGCCCT GCACTTCCTC AGCCCTAGCC
9201 GGCCCCATAA CCACCCAAGA ACAGTTATCT AGGCATCCTT CCTGAAAAGG
9251 ACTCTGAGC CTCCAGCTCA GGGGTAGAC ATATCTGGAG GCTTCTGCCC
9301 ATCCCATCTG CCCCTTCCAG GGAAAGTCCA AGTTGTTGCC TGAGAAATCA
9351 AGGGGTGCCC AGTTCTCAGC CCCCATTAGA GCAGAGTGAA CAGGGTCCCA
9401 GGTCAGGGGC TAAGAGTGCA AAGGGTTAGC CCCAACTGCT GTCCTATTCC

FIGURE 3, page 3 of 7

9451 AAGACCCTTT ACCAAAGGTG AGATCCCAGA GCTGGGAGCT AACTGGGGCA
9501 GAAACCCTGG CCCCAGGCCA ATCACACCTG CCTGCAGTCC CTTGGGCCAC
9551 CAGCAGAGGG CAGGCAACGC CTGCTTCTGG GGCAAAATAT GGGCCCGCTG
9601 GGGCGGAGGC CTCCTTCCCC AGAGTGACCC ATTTGGGCTT GACAGGCGGA
9651 TATGAGCAGC GAAAGAGTGA TCCGTTCCGA AGACGGAGAG ACCTTGGCCA
9701 GGGTAAGTGA TTGTCTGTGG GACAGGGTGA AGGGTGGGG CAACCCGACG
9751 CTGGCCCTGA GGACACTCTC TCCCGGGCAG GAGTACGGTG TTCCCTTCCT
9801 GGAGACCAGC GCCAAGACTG GCATGAATGT GGAGTTAGCC TTTCTGGCCA
9851 TCGCCAAGTG AGAGCTGGGC AGGGAAGGGA AGTGTGCGGG GCAGGGCGGC
9901 ACACCTCCAG AATCCAGTAG GGCCCGGCC CTGGCCAGC CCCTGGACAC
9951 ACCTGCATTC TGCAGGCTGA GGTCCATTTG CTCTGGGAGC ACTGGGCCAC
10001 TGGGAGAGGG GAGGGGGCGG CTCAGCTCCT CACCCAGCC CAGCCAGCC
10051 CAGCCAGCC CATTGTCTCT TCTCAAGGG AACTGAAATA CCGGGCCGGG
10101 CATCAGGCGG ATGAGCCAG CTTCAGATC CGAGACTATG TAGAGTCCCA
10151 GAAGAAGCGC TCCAGCTGCT GTCCTTCAT GTGAATCCCA GGGGCGAGAG
10201 AGGAGGCTCT GGAGGCACAC AGGATGCAGC CTTCCCTCTC CCAGGCCTGG
10251 CTTATTCCAA GAGGCTGAGC CAATGGGGAG AAAGATGGAG GACTCACTGC
10301 ACAGCCGCTT CCTAGCAGGG AGCTATACTC CAACTCCTAC TTGAGTTCTT
10351 GCGGTCTCCC CGCATCCACA GGGAGGGTAA AACACTTAGC TTTTATTTTA
10401 ATAGTACATA ATTTAATACC AAAAAAGCG CCTGGATCCC CAAAAACCG
10451 AGGCTGGGAG CTAGTGGCCC TTTTGCTTTC TAGGACTTGG GGGCCCGGCC
10501 CTCCCTCCTA AGCATAACAA AGGTGGTGTT GCTCCAGCTC AGCCCCAGGG
10551 GACACAGATG CACTTTGGGG GTGAGGGCAG GTAATGACTC CATCGCACCC
10601 TCAGTTTACG TGGACAGAGG CTCAGGTGAC CCCAGCCTTC ACTGTCTCCC
10651 GCTCTCCAGG AGCTTATCTT CGCCCCATCT CCCAAATAAG TGGGCCCTTG
10701 TGCTGTGAGG AAGACCAAG CCTCAGGGAA GATAAGAGAT ATGGAGATGG
10751 GAGGGGGAGG ACAAGGGGCA GAGAGTAGGG TCTAGCTGGC TATCTCTGGC
10801 CTTACTAACA CCCCCCTGGA GGCATGCCCC TTTTCTCCAG CACACAAGCA
10851 CATTGGGGCA CCTGGAATAA TTGGTTCAG GTCCTGTTC TCTGGACTTC
10901 AGATCTGGG GGAGCCCTC CCCCCCTGA ATCCCTGGCT TAGCTACCTT
10951 CCTGCCGTG CACCTAAAAA CCTCAGGTCA GAACTAGGAA AAGAGTTTTG
11001 TTTTATTTT TTTGAAATGG AGTCTCGTTC TGTGCGCCAG GCTGAGGTGC
11051 AGTAGTGCAA TCTCCGCTCA CTACAACCTC CACTCCCTGG GGCTCAAGCG
11101 ATCCTCCAC CTCAGCGGCC GAAGTAGCTG GGACTATAGG TGTGTACCAT
11151 CACACCTGGC TAATTTTGT ATTTTGTGA GACACAGGGT TTCGCCATGT
11201 TGCCCAAGCT GGTCTTGAAT TCCTGAGCTC AAGCAACCTG CCGGCCTCGG
11251 CCTCCCAAG TACTGGGATT ACACGCAGAA GGCACCATG CCAGGCTAGA
11301 TGTGTCTTAT CCAATCCTT TGGCAGGCAT GCAGCTCCAC AGGCGATTTC
11351 TTCAAGCAGC TGAAGTGTG AGCCCTCCTG GGTAAAGAGC CAGATAAGGA
11401 GAAATCCCTT TCCTAGGTTT GGAATGTGTT GTGAAAAAA AGAGAAATCC
11451 CTGGCTCCTG GAGCTGGTGG GAGACAAGAT TAAGCAAACC TCCCTGACA
11501 TGTATCCCTT TGACCCCAAG CTCTGCCTCC TCCCTGACCA CCCATGCCCT
11551 TTCCTTTAAT TTCTCAAACA GATACCAGG CCTAACTGC TTTACCTCCC
11601 CTCCTACTGA GTCAGGTAG GTGGTGGGAG GTCACCCATT TCCGAGTTAA
11651 ACCAATGCAA TATGAGTAAA ACAAAGTCAT GTGGGTATGT CTGGGGTAGA
11701 GAGAGGGGTA GCAAGTTCAT GTGTCTCCT TGGTCACATA TCTCCCAAAG
11751 CTCTGATCCC TGCCATGGGA AGTGGACAGG AAACATGAGG TCATGACCTG
11801 CAGGCATCTT TACTGCAGCT CTGCCGGCCT GGAGGGGGAG AGGGGGAGGA
11851 AGAAGTATGC GCTGCACATT TCTGAGGCTA CTGCATTTG TTTCAAGGCA
11901 GAAATCTTGC TCTGAGCAGT CAGCGGCTCC AGTTTGGGCC CGATAAGGAA
11951 GTTCTCCGTG GCCTCCCTCA GGCAGAGCAG GGAGGAGGCT GACATTGCCA
12001 GTCTCTTCTG GGGCCCAAGG CAGGTTGCAG GAGATCCAAT CCCATAGACA
12051 GCTCTGGGCC TCTTGCAATT GAGTTTTTCA GAATTAAACT GCAGTATTTT
12101 GGAAAGCACA TCCTGTCCAC TGTTTCTTTG AAGTGAGTGG GGGGGGGGGG
12151 TCTTGTGAA GGAATTGTCA TTTACTGCCA AAATCATTC ATCTCCTTC
12201 CTCAGTGTCT GTCCTCAGAT GGTGAGCTCC CCGCTCAACA GACTGTCTCC
12251 CGCCTCTGTG ACCACCTCT CTTTGGCAAG AGGGAGCTAG AAGGCTTTAC
12301 AGTCCTAATC ATTTTCTGT TGGAAAAAA AAAAAAAAC CAAGGCTCCT
12351 TTCCCTGTGG CGTGTACCCA GAGGTTGATT ACCTGAGTCT GTCTGCCTC
12401 TCCCCACCCC ACCTCCCTAG CCAAACGCTG CTGCCAAAGC CCACGCTATT
12451 GCCCTAGATG GCCTGTCTTC AGCGGGCTGC CCCTCGAGGT CCCAGGCTCT
12501 CCGCGGAGCC CTCACCTTCC CAGCAGGGAT CAGAACCTGC ACTCCTCTAT
12551 GCGAGTCTCG GGACAGCACA AAGTGGATTA GGGTTAGGGT TCCACAAAC

FIGURE 3, page 4 of 7

12601 GGAAAAATGT TATTCAAACA ACTCTGTAGG GTCCGAGGAG GCCCTCCGTC
12651 TTAATTCTCG AGACTGACCG GCCCTCGCTG CCCCAGACGG GAGCAGTTGC
12701 CCCGGCAACA GCCGCTCCCT CTCAACTGGA GCTGCACCCA GGCTTTGGCT
12751 AAAGGCTGTT AAAACGTTGG CCAGGTGCGG AGGCTCACGT CTGTAATCCC
12801 AGGGCGGATC ACCTGAGGTC AGGAGTTTGA AACCATCCTG GCCAACATGG
12851 CGAAATTTTCG TCTCTACTAA AAATACAAAA ATTAGCGGGG CGTGGTGGTG
12901 CGCGCCTGTA ACCCCAGCTG CTCGGGAGGC TGAGGCAGGG GAATCGCTTG
12951 AACCCTGGGAG GCGGAGGTTG CAGTGATCCG AGATCGCGCC ACGGCAGTCC
13001 AGCCTGGGCG ACAGAGCGAG ACTCCGTCTC AAAAAAAAAA AAAAAAGTTA
13051 GGGTCCTTTA CCCGAGGGCC GGCTTTCCTC ACTCCCCGCC ACAGGTAGGG
13101 GAAACCAGGC CGGAGCCGGC GGGCCACCC GCCCAGAACC GGAATTCGG
13151 CGAGCCCCGC CCCTGCCACC CCAGCGCCGG CC (SEQ ID NO:3)

FEATURES:

Start: 2042
Exon: 2042-2134
Intron: 2135-5539
Exon: 5540-5650
Intron: 5651-7026
Exon: 7027-7068
Intron: 7069-7901
Exon: 7902-7968
Intron: 7969-8060
Exon: 8061-8113
Intron: 8114-9080
Exon: 9081-9146
Intron: 9147-9645
Exon: 9646-9702
Intron: 9703-9780
Exon: 9781-9857
Intron: 9858-10078
Exon: 10079-10181
Stop: 10182

CHROMOSOME MAP POSITION:

Chromosome # 17

ALLELIC VARIANTS (SNPs):

DNA Position	Major	Minor	Domain
4259	C	T	Intron
4325	G	T	Intron
4348	G	A	Intron
4924	G	A	Intron
4983	-	A	Intron
6710	A	G	Intron
8624	A	G	Intron
8661	G	A	Intron
11754	T	C	Beyond ORF(3')
11836	A	G	Beyond ORF(3')

Context:

DNA
Position
4259

ACCCATTAAGCCACCTAACCAGCAGCTGGGAAATTCCAGCATTGGATCTAGACCCCTGTT
ATCCAAGATTGGAGAACAGTGGGACAAAGTGCTCCTCTCCACCATTCTGCGTGTCCCTG
GGGAAGATGAGCAGAGCAGAGCCAGACAGTAAAGGAGAGGGCCACGCCCCCTCCACAGGT
TACCTCCTTGGTACTCCTGCCCCGCACTACCCACAGCAACCCCGGGATGCCGATCTGCAGC
CACATGTCCCATGTGGGAGGTTTCTGCTGAAAGAACTTCCAACCTACACATCTCCCCACTT
[C, T]
AGTATAAATTTCAACCTTCCCTAATTCATGCAACCTTTTTTTTTTTTTTTTTTTTTTTGAG
ACAGAGTGTGCTCTGTACCCGAGGCTGGAGTTCAGTGATGCAATCTCGGCTCACTGCAA
CCTCTACCTCCTGGGTTCAGCTATTCTCCTGTCTCCGCCTCCCAAGTAACGGGACTAC
AGGCGTGTGCCACCACTCCTGGCTAGTTTTTTGTATTTTGTAGAGATGGGGTTTCACC
TTGTTGGTCAGGCTGGTCTCAAACCTCCCAACTCAGGTGATCCGTCCACTTGGGCACCCAA

4325 GATTGGAGAACAGTGGGACAAAGTGCTCCTCTCCACCATTCTGCGTGTCCCTGGGGAAG
ATGAGCAGAGCAGAGCCAGACAGTAAAGGAGAGGGCCACGCCCCCTCCACAGGTTACCTC
CTTGGTACTCCTGCCCCGCACTACCCACAGCAACCCCGGGATGCCGATCTGCAGCCACATG
TCCCATGTGGGAGGTTTCTGCTGAAAGAACTTCCAACCTACACATCTCCCCACTTCAGTAT
AAATTTCAACCTTCCCTAATTCATGCAACCTTTTTTTTTTTTTTTTTTTTTTTGAGACAGA
[G, T]
TGTCGCTCTGTACCCGAGGCTGGAGTTCAGTGATGCAATCTCGGCTCACTGCAACCTCTA
CCTCCTGGGTTCAAGCTATTCTCCTGTCTCCGCCTCCCAAGTAACGGGACTACAGGCGT
GTGCCACCACTCCTGGCTAGTTTTTTGTATTTTGTAGAGATGGGGTTTCACCTTGTG
GTCAGGCTGGTCTCAAACCTCCCAACTCAGGTGATCCGTCCACTTGGGCACCCAAAATG

4348 TGCTCCTCTCCACCATTCTGCGTGTCCCTGGGGAAGATGAGCAGAGCAGAGCCAGACAG
TAAAGGAGAGGGCCACGCCCCCTCCACAGGTTACCTCCTTGGTACTCCTGCCCCGCACTAC
CCACAGCAACCCCGGGATGCCGATCTGCAGCCACATGTCCCATGTGGGAGGTTCTGCTG
AAAGAACTTCCAACCTACACATCTCCCCACTTCAGTATAAATTTCAACCTTCCCTAATTCA
TGCAACCTTTTTTTTTTTTTTTTTTTTTTTTGGAGACAGAGTGTGCTCTGTACCCGAGGCTG
[G, A]
AGTTCAGTGATGCAATCTCGGCTCACTGCAACCTCTACCTCCTGGGTTCAAGCTATTCTC
CTGTCTCCGCCTCCCAAGTAACGGGACTACAGGCGTGTGCCACCACTCCTGGCTAGTTT
TTTGTATTTTGTAGAGATGGGGTTTACCTTGTGTCAGGCTGGTCTCAAACCTCCCA
ACTCAGGTGATCCGTCCACTTGGGCACCCAAAATG

4924 TTCAAGTACCAGCCTGGCCAACATGGTAGAAACCCCGTCTCTACTAAAAATAAAAAATTA
GCCAGGCGAGGTGGTGCATGCCTATAATCCCAGCTACTCAGGTAGGCTGAGGCAGGAGAA
TCATTTAAACCTGGGAGGTGGAGGTTGTGGTGAGCCAAGATCTCGCCATTGCACTCCAGC
CTGGGCAACAAGAGCAAACTCC
[G, A]
TCTCAAAAAAAAAAAGAAAGAAAGAAAGAAAGAAACTTCCAAATAAATGTTGTGACACAA
AAAAAAAAAACCCAAACAATATTATTATAGAGTATGCAATGACCATGCCCCACCCCAAG
CAGATTCTGATAGACTCCCTTGGGTGGGAATCCTTGTCCAATATATTGACACTTCCCTTT
CCTGTGAGTATAGCCAGCCCATGCGTGTACTCACGAGCGGACGATGGATGACACAAGTA
CACAGAGGGACGGAATCCCTGCATGGTGTGGCTATGGGCAATGTGGCCACTGTCTAGAT

4983 TTCAAGTACCAGCCTGGCCAACATGGTAGAAACCCCGTCTCTACTAAAAATAAAAAATTA
GCCAGGCGAGGTGGTGCATGCCTATAATCCCAGCTACTCAGGTAGGCTGAGGCAGGAGAA
TCATTTAAACCTGGGAGGTGGAGGTTGTGGTGAGCCAAGATCTCGCCATTGCACTCCAGC
CTGGGCAACAAGAGCAAACTCCGTCTCAAAAAAAAAAAGAAAGAAAGAAAGAAAGAAAC
TTCCAAATAAATGTTGTGACAC
[-, A]
AAAAAAAAAACCCAAACAATATTATTATAGAGTATGCAATGACCATGCCCCACCCCA
GCAGATTCTGATAGACTCCCTTGGGTGGGAATCCTTGTCCAATATATTGACACTTCCCTT
TCCTGTGATATAGCCAGCCCATGCGTGTACTCACGAGCGGACGATGGATGACACAAGT
ACACAGAGGGACGGAATCCCTGCATGGTGTGGCTATGGGCAATGTGGCCACTGTCTAGA
TTGTGCAATGTGGTGGTTCTCTGGGGCCACAGAGCACACTTGGGGACCTGTTTATGGT

6710 CACCAGGGACCCTCCCTCTCACCTTGACGACTCCATCTTACAAATCTGCATCAGGGATGC
TAGACGCTGCACACCTGAAGTGTTCATAGAGAAAAGTCTCACCCTGGCAGGTGGGGCT

FIGURE 3, page 6 of 7

CTACAGCTTCAAGCAGGCAGAAAGCGAACACTTCCTTCACTAGAGAATTAGTGGGCAGCT
AAAGAAAAGGTGCTGCTGCAGATGTAGCCTCAGGTCCCCAGGATGCAGGCAACACCCCA
TCTCCAGGGGCTCGGTACAGTCCCAAGGCTAGGCTCCAGGAGAGGGAGACCGAAGTGGG
[A, G]
AAAGGGCAGGGCCTCCAGCAGCAACCAGCCCTCCAGCCCTGGGCTGCCTGATCCCTGGAG
AGAGCCAGGATGTTTCTCAGGCTCCTCTTGCCCTGCTGTTGTGAGAAGGCAGTTACAGTC
CTCAGAAGGGACGACTCCACAGTGGAGGTGTCTGGGTATGGGGTTCCTGCTGCCCTGATG
GTATGATCTGGCTGGAGACGGTTCTGGGGCTCACTGCACCCACTCTAGGCCTGGAGAGGG
ACAAGAGAGGACGTCTGCAGAGCTGAGGAGCCACATGACTCCTGCCCTCCCATCCTCTG

8624 GTGCCACCTGCATAGCCCTCACTGTGATTACGAGTGTGTTTCGTGACAAAGTGTTTCAGA
ACAGCCCCCACTCCACCCTGGATAATTATCCACAGAGACCAAGGGAAAAACACAACCAGA
AAAGTCCACACATACATCCAGGGCAAGTTGCAAGAAAGTGACTCAGTCAGACAGAGTGAG
TGGTTGTATCCTCACAACCAAACTATTATAGAGACAAAATTTGATAAATTCAGCACCA
ATTTTGTTCACGACATTGTATAGGTTTCATGAATCCCCTGACCTCAAGGACAGTTTGCTG
[A, G]
TAAGCAAACCTAGGAGAATAAAACGTTTATATAGAAAAGAGGAAAATCCATGGCACTCATAC
TCCTACCTCCAACCCCATGCTCATGGCAGACATCACTAATCAATCACAGTACTTTTGATC
ACTGAAACCCCTTATGTGGTCTTAGAATCTTTAACAGGACACTCCAAGAAATCACTGCTGA
CAGCCAACCTGATTTGTGAGATAAGGTCTCCATGCATCTGGATCTCCATAGAACTGATAG
TTGCACAGCATAAAATGGTGAGGGTGGGGCCATTGTGGGTTGAGCCACCAAGGAAGGCCA

8661 TGTTCGTGACAAAGTGTTTCAGAACAGCCCCCACTCCACCCTGGATAATTATCCACAGAG
ACCAAGGGAAAAACACAACCAGAAAAGTCCACACATACATCCAGGGCAAGTTGCAAGAAA
GTGACTCAGTCAGACAGAGTGAGTGGTTGTATCCTCACAACCAAACTATTATAGAGACAA
AAATTTGATAAATTCAGCACCAATTTTGTTCACGACATTGTATAGGTTTCATGAATCCC
CTGACCTCAAGGACAGTTTGTGATAAGCAAACCTAGGAGAATAAAACGTTTATATAGAAA
[G, A]
AGGAAAATCCATGGCACTCATACTCCTACCTCCAACCCCATGCTCATGGCAGACATCACT
AATCAATCACAGTACTTTTGATCACTGAAACCCCTTATGTGGTCTTAGAATCTTTAACAGG
ACACTCCAAGAAATCACTGCTGACAGCCAACTGATTTGTGAGATAAGGTCTCCATGCATC
TGGATCTTCCATAGAACTGATAGTTGCACAGCATAAAATGGTGAGGGTGGGGCCATTGTG
GGTTGAGCCACCAAGGAAGGCCATCCAGGCCTGGATGGGCCAGAACAAAGGTACAGATGA

11754 GCTCCTGGAGCTGGTGGGAGACAAAGATTAAGCAAACCTCCCCTGACATGTATCCCTTTGA
CCCCAAGCTCTGCCTCCTCCCTGACCACCCATGCCCTTTCTTTAACTTCTCAAACAGAT
ACCAGGGCCTAAACTGCTTTACCTCCCCTCCTACTGAGTCAGGTTAGGTGGTGGGAGGTC
ACCCATTTCCGAGTTAAACCAATGCAATATGAGTAAAACAAAGTCATGTGGGTATGTCTG
GGGTAGAGAGAGGGGTAGCAAGTTCATGTGTCTCCTTGGTTCACATATCTCCCAAAGCTC
[T, C]
GATCCCTGCCATGGGAAGTGGACAGGAAACATGAGGTCTGACCTGCAGGCATCTTTACT
GCAGCTCTGCCGGCCTGGAGGGGGAGAGGGGGAGGAAGAAGTATGCGCTGCACATTTCTG
AGGCTACTGCATTTGCTTTCAAGGCAGAAATCTTGCTCTGAGCAGTCAGCGGCTCCAGTT
TGGGCCCCGATAAGGAAGTTCTCCGTGGCCTCCCTCAGGCAGAGCAGGGAGGAGGCTGACA
TTGCCAGTCTCTTCTGGGGCCCCAAGGCAGGTTGCAGGAGATCCAATCCCATAGACAGCTC

11836 GACCACCCATGCCCTTTCTTTAACTTCTCAAACAGATACCAGGGCCTAAACTGCTTTAC
CTCCCTCCTACTGAGTCAGGTTAGGTGGTGGGAGGTCACCCATTTCCGAGTTAAACCAA
TGCAATATGAGTAAAACAAAGTCATGTGGGTATGTCTGGGGTAGAGAGAGGGGTAGCAAG
TTCATGTGTCTCCTTGGTTCACATATCTCCCAAAGCTCTGATCCCTGCCATGGGAAGTGG
ACAGGAAACATGAGGTCTGACCTGCAGGCATCTTTACTGCAGCTCTGCCGGCCTGGAGG
[A, G]
GGAGAGGGGGAGGAAGAAGTATGCGCTGCACATTTCTGAGGCTACTGCATTTGCTTTCAA
GGCAGAAATCTTGCTCTGAGCAGTCAGCGGCTCCAGTTTGGGCCCCGATAAGGAAGTTCTC
CGTGGCCTCCCTCAGGCAGAGCAGGGAGGAGGCTGACATTGCCAGTCTCTTCTGGGGCCC
AAGGCAGGTTGCAGGAGATCCAATCCCATAGACAGCTCTGGGCTCTTGCAATTGAGTTT
TTCAGAAATAAACTGCAGTATTTTGGAAAGCACATCCTGTCCACTGTTTCTTTGAAGTGA